DOCUMENT RESUME

ED 310 652 HE 022 712

AUTHOR Ewell, Peter T.
TITLE Program Reviews, Inputs and Outcomes NCHEMS Monograph 5.
INSTITUTION National Center for Higher Education Management Systems, Boulder, Colo.
PUB DATE 83
NOTE 28p.
AVAILABLE FROM NCHEMS Publications, P.O. Drawer P, Boulder, CO 80302-9752 ($7.00).
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Administration; *College Planning; Educational Objectives; *Evaluation Methods; Higher Education; *Institutional Evaluation; *Outcomes of Education; *Program Evaluation; *Self Evaluation (Groups)

ABSTRACT

The characteristics of an effective program review, and the different ways in which a program review can go wrong, are discussed. The three characteristics of an effective program review are that it is systematic, regular, and comprehensive. The dimensions of each of these characteristics are discussed, as is the "art" of program review, a process that involves judgments about available data. Program reviews can go wrong in three ways: in the use of data; in the conduct of reviews; and in whether, and how, the results are used. The fallacies involved in each of these three areas are discussed, for example, the "single indicator fallacy," the "perfect data fallacy," and several kinds of false assumptions and comparisons. Also discussed is the question of how a program review can focus on educational outcomes rather than more traditional measures of inputs and efficiency. Three dimensions of the assessment of outcomes are explored: knowledge, skills, and attitudes. The primary value of an outcomes approach is considered to be its focus on the goals of higher education. (KM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCHEMS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Program Reviews, Inputs and Outputs

By Peter T. Ewell

1983

NCHEMS MONOGRAPH #5

National Center for Higher Education Management Systems
P.O. Drawer P Boulder, CO 80302
An Affirmative Action/Equal Opportunity Employer

Monographs

NCHEMS monographs are directed primarily toward administrators of higher education; and they are useful for researchers in higher education, as well. The monographs are informative studies of a variety of problems and issues that confront college and university administrators, especially in these times of dwindling enrollments and resources. The topics range from how to manage the internal processes of institutions of higher education to how to improve the outcomes of colleges and universities. While the monographs are based on careful research, they offer practical advice and solutions that are relevant for different types and sizes of colleges and universities.

The Link Between Planning and Budgeting (1981)
By Ellen Earle Chaffee

It is difficult for administrators to link planning and budgeting under the financial stringency now faced by most institutions of higher education. This monograph notes how solutions prescribed by theory do not work in higher education. Four characteristics of an optimal solution to linking planning to budgeting are proposed.
2BA379 **$7.00**

On Deciding How to Decide: To Centralize or Decentralize (1981)
By Ellen Earle Chaffee

Suppose the university must for the first time make drastic budget cuts. How should the process for distributing the reductions be defined? Credibility for a decision can be enhanced when those affected by it trust the decisionmaking process. This monograph suggests a six-step decisionmaking process to match information, expertise, values, and concern for people who must live with the decision.
2BA380 **$7.00**

Management Fads in Higher Education (1981)
By Richard Allen and Ellen Earle Chaffee

This monograph examines three popular management innovations that might be fads: (1) program budgeting, (2) costing, and (3) strategic planning. The origin and characteristics of each innovation are described, and the reasons why they became popular are analyzed. A number of potential pitfalls for administrators to avoid when using these management techniques are suggested.
2BA381 **$7.00**

Promoting the Effective Use of Information in Decisionmaking (1984)
By Peter T. Ewell and Ellen Earle Chaffee

Case studies drawn from different types of institutions illustrate how information is used for various purposes and with different outcomes depending upon the decisionmaking setting in which it is used. An alternative to traditional models of decisionmaking is proposed—"multiple advocacy"—in which superior decisions result from adopting a conflicting or dialectical decisionmaking process.
2BA382 **$7.00**

Program Reviews, Inputs, and Outcomes (1983)
By Peter T. Ewell

This monograph shows how program reviews can become an integral part of institutional decisionmaking. Some of the ingredients of an effective review process are discussed from both a conceptual and data-gathering perspective, as are typical problems encountered in designing and conducting program reviews.
2BA383 **$7.00**

Transformation Leadership for Improving Student Outcomes (1985)
By Peter T. Ewell

This monograph addresses the need for improvement in undergraduate general education, as well as the need for colleges and universities to test student knowledge and ability on a systematic basis. Four obstacles to improved undergraduate effectiveness are pinpointed. Several proven levers which are available to academic leaders to use to implement a campuswide instructional improvement program are then noted.
2BA384 **$7.00**

Recruitment, Retention, and Student Flow: A Comprehensive Approach to Enrollment Management Research (1985)
By Peter T. Ewell

This monograph proposes a model to guide a comprehensive institutional research program designed to inform enrollment management decisionmaking. It examines the design requirements for a research program, illustrates how to model longitudinal student flow, and discusses the determination of enrollment structure. Case studies provide illustrations of the proposed model.
2BA385 **$7.00**

The Costs of Assessment (1985)
By Peter T. Ewell and Dennis P. Jones

This monograph examines the direct costs of establishing an institutional assessment program as called for in recent national reports. A number of different examples are presented. Estimates of typical incremental costs for establishing and maintaining assessment programs are provided, including costs of test instruments, administration, analysis, and coordination.
2BA386 **$7.00**

4

Summary

Program review has become an important part of the planning process in higher education. Two nationwide trends indicate this: the increasing number of program reviews being carried out at colleges and universities across the country and review processes that focus on the outcomes of higher education. This monograph discusses the characteristics of an effective program review and the different ways that a program review can go wrong.

There are three characteristics of an effective program review. It is systematic, regular, and comprehensive. After noting the various dimensions of each of these aspects of a program review, this monograph discusses "the art of program review," a process that involves judgments about available data.

Program reviews can go wrong in three different ways. There can be problems that have to do with (1) the use of data, (2) the conduct of reviews, and (3) whether or not and how the review results get used. These involve a number of fallacies, for example, the "single indicator fallacy" and the "perfect data fallacy," as well as several different kinds of false assumptions and comparisons.

How does a program review focus on educational outcomes rather than more traditional measures of inputs and efficiency? The last section of this

monograph addresses three dimensions of assessment of outcomes: How do we assess value-added in knowledge, skills, and attitudes?

The primary value of an outcomes approach is that it focuses on the goals of higher education. Program reviews that are oriented towards outcomes demonstrate that the goals of higher education are indeed being met.

Program Reviews, Inputs and Outputs

By Peter T. Ewell

Introduction

Planning and program reviewers used to be a bit like "voices crying in the wilderness." Today, however, planning and programming are considered to be necessities, not luxuries. Previously, there were sufficient resources to carry out planning and program reviews, but now the demand is to "do more with less."

The universe of higher education has changed. This is due to the increasing numbers of people attending college today and several thousand new institutions that did not exist 20 years ago (community colleges and teachers' colleges that are now regional comprehensive universities). As a result, there has been a lack of agreement about what the ends of higher education should be. This suggests the need for not only planning and assessment but also a focus on educational outcomes.

It is difficult to open a copy of the Chronicle of Higher Education without seeing something new about accountability or assessment of educational outcomes, particularly at the secondary-school level. Several studies--the Nation at Risk study,

the recently released National Science Board Report, the Carnegie Commission Report, and the College Board's report on academic preparation for college--are examples of significant studies compiled in the last year. The movement towards accountability in higher education has been slower and less spectacular than in elementary and secondary education, but it has not been less present. Program review has become a very strong part of this movement.

Trends in Program Review

There are two nationwide trends in program review that set a context for discussion. The first is an increase in the number of program reviews being carried out at colleges and universities. About a third of all institutions of higher education now have formal program reviews. Additionally, more than two-thirds of the state-coordinating boards have some kind of program-review process in place. More than half of these have been developed since 1977, primarily because of decreases in public funds and, consequently, new demands for accountability.

The second nationwide trend is review processes that focus on the outcomes of higher education. Outcomes are the results of programs. This kind of issue is also being posed in the accreditation process. The Southern Association of Colleges and Schools, for example, is currently considering a set of explicit outcomes criteria for accreditation. A number of other accreditation bodies are considering doing the same. This focus on outcomes contrasts to typical program reviews of the '60s and '70s that were largely based on resources. They concentrated upon quality of faculty, peer review, quality of equipment, and so on. Although these factors are important parts of any review, the appropriate focus is the bottom line: What has the program produced?

Both trends in program review are valuable. It is hard to dispute the logic that holds that higher education, particularly public higher education,

should be held accountable for what it does, and that the product or outcome should be the standard of accountability. Measuring outcomes could be very difficult; but it should be possible, at least at the conceptual level. There is a sense, however, that the current situation is a mixed blessing, particularly when accountability issues are raised by those outside the community of higher education. It is problematical when institutions of higher education are held accountable by external agencies because external agencies tend to focus on the wrong things. Planning and program reviews can go wrong in at least 10 different ways, and should therefore be integral parts of institutional self-consciousness rather than reactions to outside pressure.

The Relationship Between Planning and Program Review

Planning and program review are activities that involve the institution as a whole. A focus on outcomes automatically moves the researchers to the level of the institution as a whole because it is extremely difficult to attribute the outcomes of an educational process to one part of the institution. A focus on educational outcomes is thus part of "institutional self-consciousness"--awareness of the institution as a whole that encourages comprehensive assessment of the functions and processes of the institution.

Program review is usually the first and often the most important part of the planning process. Through program review, an institution often finds itself "backing into" a planning process. Typically, for example, program reviews conclude that programs are not articulated as well as they should be, nor are program goals developed and communicated as well as they could be. This is the type of information that emerges gradually through assessment rather than directly through planning.

The Components of an Effective Review Process

The basic attributes of a good review process can be summed up in three words. The process should be systematic, regular, and comprehensive. A program review is a formal and systematic process that examines programs periodically over time, checks for changes, and matches program characteristics with ongoing program needs. Standards should be uniformly applied. Each program should have the chance to make its own case and be subject to essentially the same kind of justice. In this sense, the activities of departmental curricular committees are not program reviews because the same kinds of standards are not applied across all programs.

Changes in program characteristics and performance must be evaluated across time. The best statistics are therefore the ones that capture trends and detect broad patterns of performance. Program reviews should also be comprehensive in the sense that the review should cover all aspects of a program. Program review should be applied to all programs. Everybody has something to learn from this process, and should therefore be subject to it.

The model of how to run a program (see figure 1) is not very complicated, but it illustrates some of the weaknesses of program planning and assessment. This model breaks the management process into four basic phases. The textbook way to present the model is to start in the lower left-hand corner. Begin with the goals and objectives of the program. These drive a resource commitment that results in a budget. This, in turn, compels expenditures on a particular set of programs. Operating these programs over time results in either what was planned or what happens despite plans.

This model suggests that there are four different dimensions upon which to base reviews of the management of programs and institutions. The first

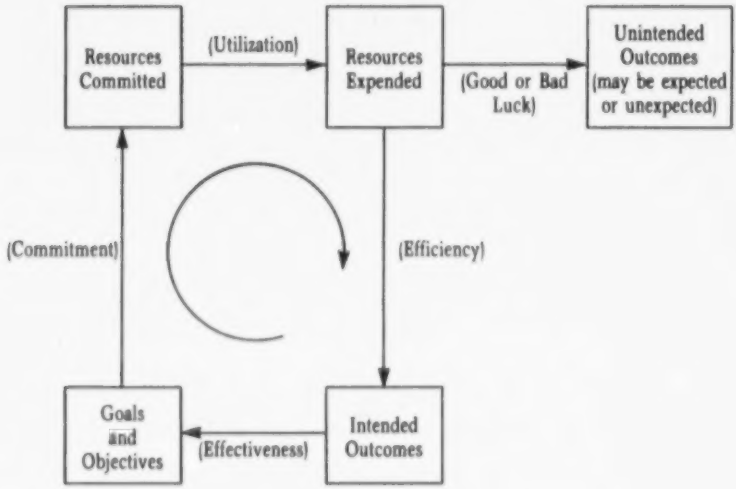

Figure 1. A Conceptual View of the Management Process

Expanded version of a framework developed by Leonard C. Romney, Robert F. Gray, and H. Kent Weldon. "Departmental Productivity: A Conceptual Framework." (Unpublished paper) Boulder, Colo.: NCHEMS, May 1978.

is efficiency, the focus of formal program reviews. This is where most program reviews get stuck. They tend to ask the question, How well utilized are the resources that have been allocated? This is a reasonable approach. It is the kind of approach that is discussed in terms of dollars expended per student credit hour, faculty contact hour loads, and--with respect to nonacademic programs--different kinds of activity measures such as numbers of square feet cleaned, numbers of crimes prevented, and so on. These are all output measures, and they are indeed important.

Processes should be designed so that they capture other changes as well, such as effectiveness. Granted, the institution has so many student credit hours coming out of a given program. But what was done with them? What happened to people who were in the program? What were the results of the program? And to what extent did the results match up with what was being attempted

to be accomplished? In other words, knowing that a program is efficient says very little about it.

On the other side of efficiency is the issue of defining the goal of the program. To what extent is the goal congruent with what it should be, given the mission of the institution and the needs of the region? Issues of efficiency, effectiveness, and program need produce a circular kind of assessment process where different kinds of data can reinforce one another. Program efficiency, the focus of most data gathering, is only a small part of the assessment process.

There are five distinct dimensions of assessment in a program review, and not all are directly measurable. Some common indicators are often measured, but these are only indicators and should be treated with caution. There is a strong tendency to think that, because there is data about a program, something is actually known about it. One of the most serious problems with program-review efforts is letting the available data define what is being evaluated, rather than the reverse. Review processes should be built from the top down. They should start out with this proposition: "Let's leave data aside for a minute and identify the dimensions along which we should be investigating programs. Then, let's try to figure out some way to measure them." The measurements may be unsatisfactory; they may be "first cuts." But it is better to have bad data about the right question than the other way around.

All program review processes should have data about five dimensions. The first dimension is level of productivity and output. The second dimension is program need. The third dimension is "characteristics of students enrolled in the program." What are students like? How do we recognize them when we see them? Does the program have many different constituencies or only one? What do they want? The fourth dimension about which there should be data is student outcomes such as

Figure 2. Productivity/Output Indicators

Program Enrollments	Five-Year Credit Hour Enrollments
	Five-Year Headcount Enrollments
Program Completion	Number of Degrees/Certificates Granted Over Five Years
	Proportion of Program Starters Completing Over (x) Years
Faculty Productivity	FTE Student/Fte Faculty Ratio
	Faculty Contacts Hour Load
Program Costs	Discipline Costs/State Average for Discipline
	Program Costs/Institutional Average

achievement. What are the results of the program? This often takes the form of a "value-added" assessment process. This is a process that is based upon the notion that education is a kind of a production process: start with an unfinished product, do something to it, and out comes a more finished product, presumably one of greater value. The entire notion of assessing student outcomes is a value-added notion.

Figure 3. Program Need Indicators

Local/Regional Job Market Needs	Number of Existing/Projected Job Openings
	Proportion of Area Employers Expressing Need for Program
Support of Academic/Occupational Area for Which Instruction is Provided	Participation in Program Advisory Committee Meeting
	Proportion of Credits Accepted in Field at Other Institutions
Success in Meeting Expressed Student Needs/Demands	Proportion of Completers/Leavers "Satisfied" with Instruction in Program
Program Autonomy	Proportion of Total Program Credit Hours Taken by Program Faculty
	Service Instruction to Other Programs
Lack of Unnecessary Duplication With Other Programs	Listing/Enrollments of Similar Programs in (Region)
	Number (Listing) of Other Courses in Same Discipline Taught at Institution

Figure 4. Student Characteristics Indicators

Student Demographics	Sex Age Race/Ethnicity Residence
Student Preparation/Ability	High School G.P.A. Test Scores Proportion Transfer
Student Activity Levels	Average Load Term Proportion Full-Time Proportion Evening
Student Attitudes/Aspirations	Occupational Aspirations Highest Degree Planned Primary Reason for Enrollment in Program

Figure 5. Student Achievement (Outcome) Indicators

Attrition/Retention	Program Completion Rate Program "Stop-Out" Rate 1st-Term Attrition Reasons for Non-Completion
Academic Achievement	Comparative Scoring on Achievement/Certification Examinations Course Completion Rates "Gain" Scores (Valued-Added) Self-Assessments of Growth in Identified Areas
Occupational/Post-Graduate Success	Proportion of Completers/Leavers Placed in Field Proportion of Completers/Leavers Accepted in Graduate/Senior Institutions G.P.A. in Senior/Graduate Programs Self-Assessments of Contribution of Instruction to —Further Education —Job Performance Annual Incomes of Completers/Leavers Working Full-Time in Field

Finally, there is the elusive dimension of quality. Quality is usually a catch-all kind of criteria. This is the dimension that captures notions of faculty quality, support services, and that unknown aura that surrounds certain programs once they have established a critical mass and a reputation. In fact, most of the indicators that are generally listed under quality could probably be put in one of the other four categories if they could be specified more clearly. It makes sense to have quality be a separate dimension, however, because it emphasizes the importance of subjective judgment. And, because the whole process is judgmental, there is no formula for carrying out a program review. It is a very human process, and it should be kept that way!

Figure 6. Program Quality Indicators

Faculty Quality	Degree Levels/Training of Full-Time Faculty
	Years of Teaching Experience of Full-Time Faculty
	Proportion of Courses Taught by Part-Time Faculty
Curriculum/Instructional Environment	Average Class Size
	Actual/Possible Course-Taking Sequences and Patterns
	Comparison with Similar "Peer" Programs
	Comparison with Established or Similar Accreditation Standards
	Student (Graduate) Assessment of Program "Quality"

The Art of Program Review

Having data about the five dimensions is different than making judgments about them. Given certain data, there are many ways of arriving at judgments. This is the art of program review.

One of the most common ways in which program reviews can go wrong is to set up mechanistic, data-driven standards. To counteract this tendency, the pattern of review results (rather than the discrete sets of indicators of program performance) should be examined. What is the pattern of each of the

dimensions? Is it is expected? Different institutions strive for different kinds of patterns of outcomes. The outcomes of programs at a community college look very different from those of a four-year institution. And the outcomes of programs at a public, regional service institution look very different from those of a small, elite liberal arts college. One of the difficulties with most program-review methodologies is that they fail to take into account institutional differences.

An example of trying to use a pattern approach to program reviews is the idea of basing program reviews on what is called a "program portfolio" (see figure 7). A panel judges two dimensions of an institution's programs. The first is "mission centrality"--the degree to which the program is essential to the mission of the whole institution. The second is the growth rate of the relevant job market in the region. The second dimension is not the one that should be chosen in all cases. In fact, the idea behind the program-portfolio approach is to ascertain the key dimension of effectiveness and plot this together with mission centrality.

Figure 7. Ten Problems and Fallacies of Program Review
NCHEMS Kellogg Student Outcomes Project

Data-Use Problems:
- The Single Indicator Fallacy
- The "Index Impulse" Problem
- The False Comparison Problem
- The "Straight-Line" Fallacy

Problems in Conducting Reviews:
- The "Perfect Data" Fallacy
- The Interdependence Problem
- The "Self-Reference" Fallacy
- The Unknown Context Problem

Problems in Applying Review Results:
- The "Process-Betrayal" Problem
- The "Process-Isolation" Problem

Note that there can be some very effective programs that are not very central. They therefore do not belong in the institution, and a decision should be made about whether or not to continue these programs.

This approach is not a panacea. There are many ways it can go wrong. It is an alternative to one-dimensional approaches that are most common. The importance of the bottom axis of mission centrality should be emphasized. To be useful for planning, a program-review process should be institutionally focused and take into account where each program fits into the mission of the whole institution.

How Program Reviews Can Go Wrong

Arranged from the relatively simple to the relatively complex, there are three categories of problems: (1) problems in the use of data, (2) problems in the conduct of the reviews, and (3) problems that have to do with whether or not and how the review results get used.

The most common way in which program reviews go wrong is captured by the notion of "single indicator fallacy." This fallacy is based on the popular idea that a program can be judged in terms of a number, one number. Generally, the number that is chosen is a cost number or a load number, for example a cost-per-student credit hour or a student credit-hour production figure for each FTE faculty by department. Reliance on a single indicator spells trouble, as the following case illustrates.

In the state of Tennessee, a rather interesting program--performance funding--has been established. The basis of this program is a set-aside of 5 percent of the state's budget for higher education that is reallocated to institutions across the state according to a set of performance-funding criteria. In its initial phases, this reallocation process took place according to what appeared to be a

standard institutional program review. The institution submitted what was essentially a self-assessment. This was a collective judgment call, and dollars were allocated on this basis. Now, an outcomes indicator based upon improvement in the performance of students on the American College Testing Program College Outcomes Measures Project (ACT-COMP) examination is being used. Regardless of the merits of the ACT-COMP exam (and they are many), it is a single test and dollars are allocated according to the results of the tests.

A look at test scores for several institutions in Tennessee clearly indicates that it is dangerous to use one indicator to allocate dollars. In fact, it is a standard axiom regarding program assessments that it is better not to use any indicators than to use one quantitative indicator. Experience and judgment on a number of factors about a program is superior to working with one indicator. This does not say that data should not be collected. Rather, this suggests that perhaps the one biggest mistake that a program review can make is to rely on a single indicator.

The second problem is the "index impulse problem." It reflects a trend in program reviews: take all the data and boil it down to one number. If there is not a single best number to begin with, how can a composite score be meaningfully derived from factors such as student/faculty ratios and generosity of funding and equipment support? Such a process enables all the programs in the institution to be compared and ranked accordingly. This kind of process can be useful, but it can also be very misleading. Inherent in it are all the difficulties of using averages in a statistical presentation. A program that is doing more or less well on most indicators does extremely well on an index. The programs that excel in some areas do not. In fact, it is often the outliers on any given indicator that reveal the most about programs.

Some of the more effective program reviews try to capitalize on this by graphing particular indicators against one another. The result is a kind of program profile that can be compared visually with the profiles of other programs. If a program looks superior on a number of different dimensions, it can be seen immediately with this technique. A procedure such as this has some of the virtues of a single indicator, but all the information of multiple dimensions.

A third problem is false comparison. There are many ways to go wrong when directly comparing numbers to one another. Comparison is very important, nonetheless. All comparisons should be treated with some caution. Two kinds of comparison are generally made in program reviews. The first is comparison of program to program within the institution. This is a problematical comparison because there are legitimate differences, such as differences in the cost of different kinds of programs, differences in the way the instruction is organized in different programs, and so on.

The second type of comparison, increasingly used in reviews, is comparison of apparently similar programs at different institutions. How does our English program stack up against another university's English program? When making this kind of comparison, close control must be kept over what is being counted. The way things are counted at different institutions varies significantly, and therefore as gross an indicator as possible should be used when making such comparisons. It is also a good idea when making comparisons of both kinds to collect as many different measures as possible of the programs being compared to see if the measures are saying the same thing. In short, comparison is the heart of analysis, but it is critical to establish the basis for comparison and keep control of it.

The fourth problem is the "straight line fallacy." It has to do with extrapolation: today's trend (as indicated by a linear pattern) will not be the same in 1995 nor will it have the same meaning. So much of the data of program review is organized linearly and tends to be extended too far into the future. This procedure also assumes that there are linear relations imbedded in program performance. For example, it assumes that student/faculty ratios and credit-hour production are related to one another in a linear fashion when, as often as not, this is not true. An example of a linear relation that is often taken as an article of academic faith is class size. The relationship between instructional effectiveness and class size is thought to be completely linear, that is, the more students there are in a class, the more efficiency increases but educational quality decreases. This is not true. In fact, if what is happening at the margin is examined when students are added to a class, it is clear that there is an enormous flat spot in the curve of effectiveness between about 20 and 40 students. The efficiency question needs to be looked at from a marginal perspective as well as from the perspective of determining the additional cost of bringing one more student on board. Even though there are many difficulties in measuring marginal costs and effectiveness, these approaches should not be ignored simply because they are difficult to measure.

The next group of problems centers on the conduct of program review. The biggest problem can be called "the perfect data fallacy." The premise of this fallacy is that reviews cannot be carried out because there is not good data. Everyone in a program that is threatened will come up with an objection of this kind! In fact, their objections might be valid and appropriate. But bad data should not be the reason for not carrying out a review. In fact, <u>all</u> data are bad in the sense that they are a snapshot of factors (like enrollments and costs) that constantly change. This does not mean that judgments cannot be made. Dennis Jones at NCHEMS

has an aphorism about this phenomenon: "In higher education we tend to measure everything with a micrometer, mark it with chalk, and then cut it with an axe." The kinds of decisions that are made, in other words, are not in the least related to the precision with which they are measured. Most decisions could have been made on the basis of much cruder data and, in some cases, no data at all.

Even though decisions can be made based on very approximate data, one should be aware of the level of error involved in each measurement that is used. Data should be subjected to a lot of sensitivity testing. In any program-review process, the first line of defense of a program that is threatened is typically: "the data are no good." This raises the question, How bad are the data actually? In some cases, the enrollment number could be doubled and the decision would still be the same. In others, the enrollment number could be cut in half and the decision would still be the same. The point is that the data show a situation that is so far out of line that the available number is good enough for purposes of program review. Double the number, halve the number, and ask what change would be necessary to affect the decision.

Another way to counter the "perfect data fallacy" is by using multiple indicators. The more that can be found out about a program, the more confidence there can be in the conclusions, even when crude data is used. If all the indicators point in the same direction, something authentic is being captured. There may not be a perfect enrollment number, there may not be a perfect load number, there may not be a perfect program demand number, and so on. But if all the numbers point in the same direction, there is a sound basis for judging the program. And making sound judgments is the goal of program reviews.

The next problem is the problem of inter-dependence. The essence of this problem is that it is extremely difficult to treat programs as though

they were free-standing, as though they do not relate to other programs or services that the institution offers. A major component of any review should be examination of the instructional service that each program provides others through an induced course-load matrix. Frequently, a program in itself might be weak, but, when considered as a service to the institution, it is strong. Because of the interdynamics among programs, the elimination of a particular program can do more harm than good when, for example, it does not save money and results in loss of an important service. On the other hand, a program might have little or no interaction with the rest of the institution. If other indicators are also negative, there is a good case for eliminating such programs.

A related problem is the "self-reference fallacy." This is the idea that a program should be judged solely on the basis of what the program purports to accomplish. The program should also be judged in terms of what the institution as a whole is trying to accomplish. Most of the time, program review is a process that analytically breaks down the institution. Indeed, the fact that the unit of analysis of program reviews is the individual program makes it tempting to conduct program reviews as though each program were a free-standing entity with its own separate goals. One way to guard against this isolationism is through the program-portfolio approach described earlier.

A final problem is the "unknown context problem." This problem is a more general version of the "self-reference fallacy." Any review, if it is effective, should proceed in the context of known institutional constraints and priorities. When institutional constraints and priorities are out on the table at the beginning of the process, a lot of conflict can be avoided. If not, review findings can lead to simplistic notions such as, "If only we had three more faculty and $12,000 more in travel money, everything would be all right."

The point is that new resources are not available, and the program review must take into consideration from the outset the known constraints. Clear communication is a prerequisite so that the program-review process can proceed as a collective enterprise and be appropriately focused at the institutional level. The people conducting the review, including program faculty and staff members, should be aware of the institutional constraints and plan accordingly. They should be aware that it is not a simple matter of going to the legislature and saying, "We want more."

There are two final problems. These deal with how review results should be used. The first of these problems is "process betrayal." Like many planning processes, program review is often instituted for the wrong reasons. For example, administrators think that program review enables them to control an institution more closely. An attempt to use program review for control can be called "process betrayal." This happens when an administrator says, "I'm sorry we're not going to abide by this process any more because it didn't show us what we wanted to see," and walks away. This is the quickest way to ruin a planning process and ensure that no one will take it seriously for at least five years. To avoid this, program review must be thought of as a bargain. It is an agreement on the part of all the parties involved to, first, abide by the rules and, second, if the review comes out with results that recommend a particular direction, treat the recommendation seriously and implement it in the planning/budgeting process. A program might not get the requested funding, but the issues that were brought out in the review will have been seriously considered in the course of the budget process.

The best way to avoid process betrayal is to avoid "process isolation"--the notion that program review is free-standing and does not interact with planning or decisionmaking. Often program review is implemented only because it improves the quality

of individual programs. This is a good reason to carry out reviews, but a better reason is to improve overall planning and budgeting. Results of reviews should be fed back into the budget process, they should be shown in statements of institutional priorities, and they should be reflected in a wide range of administrative decisions. It is important to distinguish between a review process that is directed largely at improvement of individual programs and one that is directed at the institution as a whole. With respect to the former, the goal is a better program; with respect to the latter, the goal is to ascertain the trade-offs and priorities among all of an institutions programs. Even though there is always some tension between these two objectives, a good review process serves both.

An Outcomes Approach

How should reviews be designed so that they focus on the notion of educational outcomes rather than more traditional measures of inputs and efficiency? Under the auspices of the Kellogg Project, seven public institutions are attempting to make better use of the data that they have collected about their students. Four of the seven institutions are heavily involved in program review, and they are trying to orient their program reviews toward educational outcomes rather than efficiency. It is not easy, but it can be done.

Any institution that is serious about quality should consider three things about its students--three basic dimensions of assessment of outcomes. The first is a cognitive dimension, and asks, "Do our students know more when they are finished with their education than when they started?" There are many problems measuring this type of outcome. More important, there are questions about what should be measured. There is a strong tendency, for example, to test students on the basis of a particular, established body of knowledge. More and more, people are starting to look at assessments of broad,

liberal-education outcomes. For example, the widely used ACT-COMP examination attempts to do this, and several states have mandated its use. Liberal education is particularly hard to assess because it is usually so badly defined. As Robert Pace of the University of California at Los Angeles (UCLA) has noted, "Statements of the goals of liberal education are statements of faith and hope, and should be read with charity." Nonetheless, efforts to assess these goals should not be abandoned.

A second dimension of assessment is quite different: "Can our students do anything better as a result of instruction?" This is not a question about knowledge, it is a question about application of knowledge. Have students gained any new skills? Have they been trained effectively to do certain kinds of things? This assessment does not need to be confined to a set of specific occupational skills. For example, Bob Pace, one of today's pioneers in education, attempts to assess what can be called "quality of student effort." This consists of a set of skills that a student picks up in the course of a college education for which Pace has designed measurement instruments.

A third dimension of assessment is captured by this question: "What are a student's feelings, attitudes, and changes in values as a result of the educational experience?" How do they think they have progressed? An institution that is seriously interested in improving its quality should know something about its students with respect to each of these dimensions. This knowledge is an important part of any process of program review.

It is also important to point out that focusing on outcomes rather than resources implies a critique of the way educational institutions are typically organized and managed. We tend to look at "pieces of students." We do not look at the whole student. A financial-aid director looks at a particular "piece of a student," a faculty member looks at all the "problem pieces of a student." Everyone is

treating a particular portion of the activity, but it is no one's responsibility to put all the pieces together to form a coherent picture of what student experience looks like. Usually, there is no place in the institution for a holistic approach. This is why organizing effective student-retention programs is often so difficult. Student success is everybody's business, but no one's explicit responsibility.

The Kellogg Project tries to deal with problems by putting together "student-success committees" that cut across traditional academic and administrative boundaries. These committees often succeed in getting people to talk with one another when normally they would not. And an important basis for discussion in these committees is a body of outcomes data for the members of the committee to interpret.

There are a couple of lessons in this experience for designing good program reviews. First, the best program-review committees work the same way as student-success committees. They involve representation from many different constituencies, both because of the different perspectives involved and because different people have different kinds of information available to them. Second, it is important to have explicit outcomes criteria built into the planning and review process. It should be well-known at the institution that programs will be evaluated on the basis of their outcomes. Unit plans and budgets should therefore be built around objectives which are outcome-oriented.

Conclusion

The primary value of an outcomes approach is that it focuses on the goals of higher education. Program reviews demonstrate that higher education does in fact make a difference. It is time to broadcast this! One of the major difficulties is that people are frightened of measuring outcomes

because they are afraid of what they might find. Past claims based upon "the self-evidence of the importance of higher education" have helped neither the case for higher education as a whole nor the identification and correction of deficiencies. But time is running short. If this kind of accountability is not self-imposed in higher education in a reasonable, professional, and participatory manner, it will be imposed from outside the institution.

01830116000401
2C:386:LP:2BA383